

**1 A phylogenomic perspective on the radiation of**  
**2 ray-finned fishes based upon targeted sequencing of**  
**3 ultraconserved elements**

4 Michael E. Alfaro<sup>1,2,\*</sup>, Brant C. Faircloth<sup>1,2</sup>, Laurie Sorenson<sup>1</sup>, Francesco Santini<sup>1</sup>

5 <sup>1</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

6 <sup>2</sup>These authors contributed equally to this work

7 \* E-mail: michaelalfaro@ucla.edu

## Summary

Ray-finned fishes constitute the dominant radiation of vertebrates with over 30,000 species. Although molecular phylogenetics has begun to disentangle major evolutionary relationships within this vast section of the Tree of Life, there is no widely available approach for efficiently collecting phylogenomic data within fishes, leaving much of the enormous potential of massively parallel sequencing technologies for resolving major radiations in ray-finned fishes unrealized. Here, we provide a genomic perspective on longstanding questions regarding the diversification of major groups of ray-finned fishes through targeted enrichment of ultraconserved nuclear DNA elements (UCEs) and their flanking sequence. Our workflow efficiently and economically generates data sets that are orders of magnitude larger than those produced by traditional approaches and is well-suited to working with museum specimens. Analysis of the UCE data set recovers a well-supported phylogeny at both shallow and deep time-scales that supports a monophyletic relationship between *Amia* and *Lepisosteus* (Holostei) and reveals elopomorphs and then osteoglossomorphs to be the earliest diverging teleost lineages. Divergence time estimation based upon 14 fossil calibrations reveals that crown teleosts appeared 270 Ma at the end of the Permian and that elopomorphs, osteoglossomorphs, ostarioclupeomorphs, and euteleosts diverged from one another by 205 Ma during the Triassic. Our approach additionally reveals that sequence capture of UCE regions and their flanking sequence offers enormous potential for resolving phylogenetic relationships within ray-finned fishes.

## Introduction

The ray-finned fishes (Actinopterygii) constitute the dominant radiation of vertebrates on the planet including more than 32,000 species and equaling or exceeding richness estimates for the combined total of birds, mammals, and reptiles. Despite a long history of systematic study, resolution of phylogenetic relationships within this vast radiation remains elusive. Studies based upon traditional morphological and single-gene, PCR-based molecular approaches have succeeded in delineating several major lineages of ray-finned fishes, but conflict over how these lineages are related to one another remains. For example, the earliest morphological studies of ray-finned fishes unite gar (*Lepisosteus*) with the bowfin (*Amia*) in the clade Holostei [1, 2] though this clade is not recovered in some later analyses [3, 4]. The early branching of teleost lineages has also been historically contentious. Systematists agree on the four earliest-diverging lineages: the osteoglossomorphs (bony-tongues; arawanas, elephant fishes, and allies), the elopomorphs (tarpons, bonefishes, and eels), the ostarioclupeomorphs (anchovies and herrings, minnows, characins, catfishes, and electric eels), and the euteleosts (salmons, pikes, lizardfishes, and perch-like fishes). However, there is disagreement over both the relationships among these groups and the basal divergences within euteleosts. Recent morphological and molecular studies support a sister-group relationship between ostarioclupeomorphs and euteleosts [5–7], but beyond this there is little agreement regarding the relationship among these ancient teleost lineages. Morphological analyses alternatively place the osteoglossomorphs [8] or the elopomorphs [5, 9–11] as the sister group to all other teleosts and the remaining lineages sister to the ostarioclupeomorph/euteleost clade. Some molecular analyses place elopomorphs and osteoglossomorphs as the sister group to remaining teleosts [12, 13] while others recover a basal divergence between osteoglossomorphs and other teleosts [7, 14].

Recently, Near *et al.* [14] used wide-spread taxonomic sampling, in conjunction with sequence collected from nine commonly used nuclear genes, to provide a more comprehensive phylogenetic hypothesis of relationships among fishes. Their results supported the mono-

phyly of the Holostei, suggested that the elopomorphs formed the earliest diverging teleost lineage [15], and provided a new timescale for the divergence of ray-finned fishes. Although promising, these new insights into the radiation of actinopterygians relied upon a relatively modest number of genomic markers, and the stability and timing of these relationship encoded throughout the genomes of the target groups remain largely untested. One exception to this statement includes a recent study by Zou *et al.* [16] which used transcriptome sequences to examine basal divergences within euteleosts. However, the Zou *et al.* [16] study did not include several anciently diverging lineages (e.g. *Amia*, osteoglossomorphs) informing questions about the early evolution of major groups of ray-finned fishes.

Phylogenomics and next-generation sequencing technologies offer enormous promise for resolving relationships within actinopterygians and other major sections of the Tree of Life. However, revolutions within genomics and informatics have had a surprisingly modest effect on data collection practices within the phylogenetics community: most studies of non-model organisms continue to rely upon direct sequencing of a moderate number of loci, and workflows that do take advantage of massively parallel sequencing platforms remain bottlenecked by cross-species amplification of phylogenetically informative loci. Several alternatives to traditional phylogenetic workflows exist that help to overcome the inefficiencies of gene-based sequencing. One class of these methods is exemplified by the recent work of Zou *et al.* [16], who used a combination of *de novo* transcriptome sequencing, existing transcript data, and computational methods to identify 274 orthologous groups from which they inferred the phylogeny of the Actinopterygii. The benefits of their approach include the use of existing, transcript-related data sets (ESTs in GenBank); reasonably well-established data generation methods; and the collection of data from hundreds of loci across the genomes of the focal taxa. Limitations of this approach include reliance on sampling fresh or properly preserved tissues (generally precluding the use of thousands of existing museum samples), dependence of the approach on expression patterns of the tissue sampled, and collection of data from fewer genomic locations than alternative methodologies.

A second class of phylogenomic methods involves sequence capture of nuclear regions flanking and including ultraconserved elements (UCEs) [17]. Rather than sequencing expressed portions of the genome, the UCE-based approach involves enriching organismal DNA libraries for hundreds to thousands of UCEs and their flanking regions; sequencing these libraries using massively parallel sequencing; and assembling, aligning, and analyzing the resulting data using informatic tools. This approach has been successfully used in mammals [18], birds [17, 19], and reptiles [20] to generate phylogenomic data sets that contain at least one order of magnitude more characters than those generated using PCR and to resolve historically contentious sections of the Tree of Life [18, 20]. The UCE approach differs from transcript-based phylogenomic studies [16] because data collection is independent of expression pattern, researchers can prepare and enrich libraries from existing tissue collections, and UCE loci may be better conserved and more numerous across distantly related taxa [18].

Here, we apply the UCE approach to ray-finned fishes by developing a novel set of sequence capture probes targeting almost 500 UCE regions in ray-finned fishes. We use the UCE data to provide the first phylogenomic perspective based upon widespread sampling of hundreds of markers across the genome on long-standing controversies regarding relationships at the base of the ray-finned fish Tree of Life. These include whether *Lepisosteus* and *Amia* form a monophyletic group (the Holostei [1, 2, 21]) and how the major lineages of teleosts, which constitute >99% of ray-finned fishes, are related to one another [5–7, 9–11, 22, 23]. We also use 14 fossil calibrations to provide the first time-scale for ray-finned fishes based upon UCE regions and their flanking sequence. Our results reveal that sequence capture of UCE regions can efficiently and economically generate massive data sets with strong resolving power at both deep and shallow phylogenetic scales within fishes.

## Materials and Methods

### Identification of UCE regions

To identify ultraconserved elements (UCEs) in fishes, we used genome-to-genome alignments of stickleback (*Gasterosteus aculeatus*) to medaka (*Oryzias latipes*) to locate nuclear DNA regions of 100% conservation greater than 80 bp in length. To enable efficient capture-probe design, we buffered these regions to 180 bp (where needed) by including equal amounts of medaka sequence 5' and 3' to each UCE. We aligned or re-aligned these buffered regions to the genome-enabled fishes (zebrafish, *Danio rerio*, stickleback, medaka, and two species of puffers, *Tetraodon nigroviridis* and *Takifugu rubripes*) using LASTZ [24], keeping only non-duplicate matches of  $\geq 120$  bp and  $\geq 80\%$  sequence identity across all species in the set. Based on the intersection of UCE loci across all fishes that were greater than 10 Kbp apart, we designed a pilot set of 120 bp sequence capture probes for each of the UCEs present among all members of the set by tiling probes at 4X density. We had these probes commercially synthesized into a custom SureSelect target enrichment kit (Agilent, Inc.). We used a higher than normal [25] tiling density to help ameliorate potential sequence differences among species introduced by buffering shorter UCEs to 180 bp.

### Library preparation, UCE enrichment, sequencing, and assembly

We prepared DNA libraries from 18 fish species, including representatives of five acanthomorph orders and two families of perciforms (Table 2), by slightly modifying the Nextera library preparation protocol for solution-based target enrichment [17] and increasing the number of PCR cycles following the tagmentation reaction to 20. Following library preparation, we substituted a blocking mix of 500  $\mu$ M (each) oligos composed of the forward and reverse complements of the Nextera adapters for the Agilent-provided adapter blocking mix (Block #3). We incubated species-specific libraries with synthetic RNA probes from the SureSelect kit for 24 h at 65°C. We followed the standard SureSelect protocol to enrich DNA

libraries following hybridization; we eluted clean, enriched DNA in 30  $\mu$ L of nuclease free water; and we used 15  $\mu$ L of enriched template in a 50  $\mu$ L PCR reaction of 20 cycles combining forward, reverse, and indexing primers with Nextera polymerase to add a custom set of 24 indexing adapters [26]. PCR clean-up was completed using Agencourt AMPure XP. We quantified enriched, indexed libraries using qPCR (Kapa Biosystems), and we prepared two library pools containing 10 libraries at equimolar ratios prior to sequencing.

We sequenced each pool of enriched DNA using two lanes of a single-end 100 bp Illumina Genome Analyser (GAIIx) run. After sequencing, we trimmed adapter contamination, low quality bases, and sequences containing ambiguous base calls using a pipeline we constructed (<https://github.com/faircloth-lab/illumiprocessor>). We assembled reads, on a species-by-species basis, into contigs using Velvet [27] and VelvetOptimiser. Following assembly, we used a software package (<https://github.com/faircloth-lab/phyluce>) containing a custom Python program (match\_contigs\_to\_probes.py) integrating LASTZ [24] to align species-specific contigs to the set of probes/UCEs we used for enrichment while removing reciprocal and non-reciprocal duplicate hits from the data set. During matching, this program creates a relational database of matches to UCE loci by taxon. This program also has the ability to include UCE loci drawn from existing genome sequences, for the primary purpose of including available data from genome-enabled taxa as outgroups or to extend taxonomic sampling. We used this feature to include UCE loci we identified in the genome sequences of *Gasterosteus aculeatus*, *Haplochromis burtoni*, *Neolamprologus brichardi*, *Oreochromis niloticus*, *Oryzias latipes*, *Pundamilia nyererei*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Gadus morhua*, and *Lepisosteus oculatus*. After generating the relational database of matches to enriched sequences and genome-enabled taxa, we used additional components of PHYLUCe (get\_match\_counts.py) to query the database and generate fasta files for the UCE loci we identified across all taxa. Then, we used a custom python program (seqcap\_align\_2.py) to align contigs with MAFFT [28] and trim contigs representing UCEs, in parallel, across the selected taxa prior to phylogenetic analysis [17].

## Phylogenetic Analyses

The large number of UCE loci we collected create a vast potential space for partitioning data that makes a traditional evaluation of alternative partitioning strategies computationally intractable. As a result, we modeled nucleotide substitutions across the concatenated data set using two approaches. For Bayesian analysis, we used a custom script (`run_mraic.py`) wrapping a modified MrAIC 1.4.4 [29] to find the best-fitting model for each UCE locus, we grouped loci having similar substitution models (selected by AICc) into the same partition, and we assigned the partition specific substitution model to all loci concatenated within each partition. For maximum likelihood analyses, we maintained the partitions identified in the Bayesian analysis and we modeled each partition using the GTR+CAT approximation. We performed Bayesian analysis of the concatenated data set using MrBayes 3.1 [30] and two independent runs (4 chains each) of 5,000,000 iterations each, sampling trees every 500 iterations, to yield a total of 10,000 trees. We sampled the last 5,000 trees after checking results for convergence by visualizing the log of posterior probability within and between the independent runs for each analysis, ensuring the average standard deviation of split frequencies was  $< 0.001$ , and ensuring the potential scale reduction factor for estimated parameters was approximately 1.0. We performed maximum likelihood analysis of the concatenated data in RAxML [31] using the rapid bootstrapping algorithm and 500 bootstrap replicates.

Gene tree-species tree methods enjoy some advantages over the analysis of concatenated data sets under certain conditions [32–34] but may also be sensitive to missing data [35] and to the resolution of individual gene trees [36]. To minimize the number of unresolved gene tree topologies and maximize the number of topologies that overlapped in sampling the base of the actinopterygian tree, we selected a subset of the UCE contigs containing complete data for *Polypterus* and *Acipenser* and loci  $\geq 50$  bp, and we used this subset to estimate a species tree with CloudForest (<https://github.com/ngcrawford/CloudForest>), a parallel implementation of a workflow combining substitution model selection (identical to MrAIC 1.4.4 [29]) and genetree estimation using PhyML [37]. We estimated the species tree by



summarizing gene trees using STAR [38–40]. To assess confidence in the resulting species tree, we used CloudForest to generate 1000, multi-locus, non-parametric bootstrap replicates by resampling nucleotides within loci as well as resampling loci within the data set [41], we summarized bootstrap replicates using STAR, and we reconciled bootstrap replicates with the species tree using RAxML.

## Divergence Time Estimation

We used a set of 14 calibration points (Appendix: Fossil calibrations) from previous timetree studies [42, 43] to date several key splits on the tree ranging in age from the Givetian (392 Ma) to the early Rupelian (32 Ma). We used the RAxML topology (Fig. 1) plus fossil-derived minimum and maximum age constraints to infer divergence times in BEAST v1.72 [44] assuming a GTR model of sequence evolution with gamma-distributed rate variation. Preliminary analyses with the full UCE data (all 491 loci) showed poor mixing under a wide range of fossil constraint parameterizations, and all models incorporating the full UCE data set failed to reach convergence after 300 million generations. To estimate divergence times for these taxa from a large data set, we randomly sampled 50 UCE loci, with replacement, to create five different matrices for BEAST analyses. We analyzed the resulting matrices using independent runs of BEAST. MCMC runs of these data sets mixed much better and showed strong evidence of convergence, having ESS values  $> 200$  for nearly all parameters after 50 million generations. However, the parameters for the mean and variance of the lognormally distributed rates mixed much more slowly with autocorrelation times of  $\geq 5 \times 10^5$  generations and yielding ESS of  $\approx 50 - 100$  across the replicates.

## Results and Discussion

### Probe design, UCE enrichment, and sequencing

We located 500 UCEs shared among all actinopterygian fishes, and we designed a set of 2,000 capture probes targeting each of these loci (4X coverage). Following enrichment and sequencing, we obtained an average of 2,819,047 reads per species, which we assembled into an average of 665 contigs having an average length of 457 bp (Table 2). After removing contigs that matched no UCEs and UCE loci that matched multiple contigs, we enriched an average of 332 (50%) unique contigs matching UCE loci from each species. Average sequencing depth across UCE loci was 498X. We integrated extant genomic data from several fish species to this group of unique UCE contigs, and we constructed 491 alignments (average length: 305 bp, 95 CI:  $\pm 16.0$ ) comprised of 149,246 characters. Each alignment contained an average of 21 target taxa (95 CI  $\pm 0.4$ ) after data trimming, and we used this incomplete data matrix for subsequent analyses with RAxML and MrBayes. After removing loci having missing data for *Polypterus* and *Acipenser*, we input 136 alignments (41,731 characters; average length: 307 bp, 95 CI:  $\pm 27.7$ ) to CLOUDFOREST for model selection and subsequent species tree estimation using STAR.

### A phylogenomic perspective on the basal radiation of ray-finned fishes

Maximum likelihood analysis produced a single, completely resolved topology wherein all but two nodes received high ( $\geq 0.99$ ) bootstrap proportions and Bayesian posterior probabilities (Fig. 1). This topology provides new insight into several long-standing questions concerning the evolution of ray-finned fishes. Our analysis strongly supports the monophyly of the Holostei (*Amia* + *Lepisosteus*). This clade is historically controversial because morphological studies alternatively support [1, 2, 21] and refute [3, 4] the monophyly of this group, while recent molecular studies generally recover the relationship [15, 45, 46]. Additionally, our

analyses do not support prior findings of an “ancient fish clade” including the Holostei + Acipenseriformes as the sister group to the teleosts [46, 47]. Rather, our results strongly suggest a traditional relationship in which these lineages form successive sister groups to the teleosts.

Our phylogenomic data provide strong evidence for the placement of elopomorphs as the sister group to all other teleosts and osteoglossomorphs and ostarioclupeomorphs as successive sister lineages to the euteleosts (Fig. 1). Our maximum likelihood topology is strongly incongruent with mitogenomic studies [7, 14] but consistent with both a recent analysis of multiple nuclear genes [15] and some of the earliest morphological analyses of the group [5, 9–11]. Within euteleosts, our results are congruent with recent molecular studies [6, 15, 16] in placing esociforms as the sister to salmoniforms rather than any neoteleost lineages.

Within acanthomorphs, the largest clade of euteleosts, UCEs recover several intriguing clades that agree with results from recent molecular phylogenetic studies. These include the African cichlids + medaka (Clade C1, Fig. 1), corresponding to an expanded clade of atherinomorphs suggested by recent studies [16, 48, 49]; a clade of gasterosteiforms (stickleback) and scorpaeniforms (*Taenionotus*) that is congruent with recent molecular and morphological studies [16, 50, 51]; and a clade including surgeonfish, frogfishes, and pufferfishes (acanthuroids, lophiiforms, and tetraodontiforms) corresponding to acanthomorph clade “N” of Dettai and Lecointre [15, 52]. Together with the estimated topology, our timetree suggests that UCEs provide sufficient phylogenetic signal to resolve divergences within fishes that are as young as 5 Ma and as old as 400 Ma.

The STAR topology was less resolved than topologies based upon analysis of the concatenated data set (Fig. S1) but recovered largely congruent relationships including a monophyletic Holostei as the sister to other actinopterygians, monophyly of elopomorphs, osteoglossomorphs, ostarioclupeomorphs, and euteleosts, and a successive sister group relationship between ostarioclupeomorphs, *Salvelinus* + *Umbra*, and all remaining euteleosts.

The species tree switched the position of the Gadiformes, represented by cod (*Gadus*) and Myctophiformes, represented by *Diaphus*. This resolution is not congruent with results from Near *et al.*, [15] but has been suggested before in other molecular studies [6, 53].

## A phylogenomic timescale for the radiation of ancient fish lineages

A general expectation during Bayesian analysis is that runs will converge more quickly on the true posterior distribution with increasing data. However, in our divergence time analysis, we were only able to obtain acceptable mixing in BEAST when analyzing subsamples of the data set; analyses of all 491 loci did not converge despite a diversity of partitioning and calibration strategies combined with long MCMC analyses. This was surprising given the relatively modest number of taxa included in this study and because we fully constrained the topology. Discordant fossil calibrations [54] could potentially underlie poor mixing behavior. However, alternative fossil calibration strategies that enforced subsets of the 14 calibrations revealed that even a single constraint on the age of the root node resulted in analyses with poor mixing behavior. We observed additional evidence that fossil conflict was not the cause of poor mixing because subsampled analyses including all 14 constraints converged and produced nearly identical credible intervals for all nodes (Fig. S2). One possible barrier to good mixing during analyses of the full data set in BEAST may be that the proposal mechanisms and/or tuning parameters for branch lengths are not suited for alignments that contain both very quickly (UCE flanks) and very slowly (UCE core regions) evolving sites. With a large number of sites, even a relatively small change in branch length could result in a large difference in likelihood, resulting in a high rejection rate of branch length proposals. This explanation is consistent with the extremely low acceptance rates we observed in our analyses ( < 5% for branch rate and node height proposals). The recovery of similar age estimates across UCE subsamples (Fig. S2) suggests that the estimated time scale for ray-finned fish diversification adequately reflects the signal for divergences contained within the full UCE data set.

Our divergence time analyses suggest that the extant radiation of ray-finned fishes began  $\approx 420$  Ma near the end of the Silurian. Divergences amongst the actinopterians, neopterygians, and holosteans (Fig. S2, nodes 1-4) are very ancient and span the late Devonian to the early Permian. Extant teleosts (Fig. S2, node 5) can trace their origin back to  $\approx 270$  Ma during the late Permian with primary divergences amongst the elopomorphs, osteoglossomorphs, ostarioclupeomorphs, and euteleosts occurring before the end of the Triassic. The credible intervals surrounding the age of teleosts are wide, however, and teleosts may have originated as early as 309 Ma during the upper Carboniferous to as late as 226 Ma during the mid-Triassic. We recover a basal split between *Salvelinus* + *Umbra*, representing two protacanthopterygian lineages, and the rest of the euteleosts, at 170 Ma during the Jurassic (Fig. S2, node 9). Extant acanthomorphs, which comprise the bulk of teleost diversity including over 16,000 species and 300 families, trace their origins to the Cretaceous,  $\approx 124$  Ma (Fig. S2, node 14). We report ages for all splits in Table 1.

Our analysis represents the first timescale for ray-finned fishes derived from widespread sampling of the nuclear genome. Previous divergence time analyses based upon fossils, mitochondrial DNA, and nuclear DNA have produced conflicting timescales for ray-finned fish diversification. For example, the earliest known fossil teleosts are elopomorphs and ostariophysans, dating to the Late Jurassic [42,43]. In contrast mitogenomic data consistently suggest Paleozoic divergences, usually into the early Permian or late Carboniferous (e.g., [55]). Some divergence time studies based upon nuclear genes have posited Late Triassic to Early Jurassic divergences (173-214 Ma [42,43]) for the teleosts, while more recent studies that include additional nuclear loci and an improved set of fossil calibrations push the age of crown teleosts to 307 Ma at the end of the Carboniferous [15]. Our UCE-derived date for the divergence of teleosts ( $\approx 270$  Ma) is much older than an earlier estimate based upon *RAG1* sequence [43] and largely overlaps with the age estimate derived by Near et al [15]. Other splits within our UCE timescale generally correspond to those derived by Near et al [15] (Fig. S3). One major exception to this observation is divergence of the euteleosts

which UCE data estimate to be  $\approx 171$  Ma (95% credible interval, 152-194 Ma) while traditional nuclear loci estimate an age of  $\approx 229$  Ma (95% credible interval 220-259 Ma) [15]. One possible cause for this difference in age estimates is that we have not sampled some of the earliest diverging eutelost lineages in our study including *Lepidogalaxias*, resulting in an underestimate of the true euteleost crown age. Our molecular timescale provides additional evidence for a relatively ancient origin of teleosts and further highlights the apparent gap in the fossil record between stem and crown members of this clade [15, 42].

## Conclusions

Sequence capture of regions anchored by UCEs offers a powerful and efficient means of generating massive genomic data sets capable of resolving phylogenetic relationships at both deep and shallow scales in non-model organisms. Our UCE-based approach offers several advantages over previous studies that should contribute to the reliability of our topology. These benefits include efficient sampling of sequence data across individual genomes and among divergent taxa, collection of data from an order of magnitude more loci than studies based upon traditionally-used genetic markers and almost twice as many loci as transcriptome-based genomic studies [16], validity of the UCE probe set across bony fishes spanning 350 Ma of evolutionary history, and utility of the UCE enrichment approach with tissues collected from museum specimens. Additionally, these data illustrate that biologists can use UCE-based genetic markers to reconstruct the phylogeny of taxa other than amniotes, supporting the observation that UCE-based markers are a universal source of phylogenetically informative characters [17, 18].

## Acknowledgments

National Science Foundation grants DEB-6861953 and DEB-6701648 (to MEA) and DEB-1242260 (to BCF) provided partial support for this work. Funds from an Amazon Web

Services education grant (to BCF) supported computational portions of this work. We thank Bryan Carstens, Scott Herke, and the LSU Genomics Facility for help with Illumina sequencing. We thank Travis Glenn for helpful discussion of several laboratory methods, and we thank John Huelsenbeck, Brian Moore, and John McCormack for helpful discussions relative to phylogenetic analyses. Marc Suchard provided helpful advice about BEAST analyses. Amisha Gadani provided illustrations. We obtained tissues used in this study on loan or as gifts from Peter Wainwright (UC Davis), Rita Metha (UC Santa Cruz), Mark Westneat (Field Museum), Eric Hilton and Patrick McGrath (Virginia Institute of Marine Sciences), David Jacobs and Ryan Ellingson (UCLA), H.J. Walker and Phil Hastings (Scripps Institute of Oceanography), Anindo Choudhury (St. Norbert College).

## Author Contributions

MEA, FS, LS, and BCF designed the study. BCF identified conserved regions and designed capture probes. LS and BCF extracted DNA, prepared DNA libraries, and performed DNA sequence enrichments. BCF processed sequence data, assembled sequence data, and created data sets for subsequent analysis. BCF performed gene tree analyses. BCF and MEA performed concatenated analyses. MEA and FS performed dating analyses. MEA, BCF, and FS wrote the manuscript. All authors discussed results and commented on the manuscript.

## Data Availability

Contigs assembled from raw read data are available from Genbank (Accession #s: XXX - YYY) **[We are in the process of submitting assembled genomic data to Genbank for archiving, which takes a bit of time given the total amount of data we are submitting]**. Probe data, assembled contigs, alignments, and data sets we used for analysis are temporarily available from <http://fish-phylogenomics.s3-website-us-east-1.amazonaws.com/> (we do not log accesses of this website). Upon acceptance of this manuscript,

358 we will make all assembled contigs, capture probes, alignments, and datasets available from  
359 Dryad (<http://datadryad.org>). Protocols for library preparation and UCE enrichment  
360 are available under Creative Commons license from <http://ultraconserved.org>. Software  
361 used for the analysis of raw sequence data are available under BSD-style license from [https:](https://github.com/faircloth-lab/splitaake)  
362 [//github.com/faircloth-lab/splitaake](https://github.com/faircloth-lab/splitaake), <https://github.com/faircloth-lab/illumiprocessor>,  
363 and <https://github.com/faircloth-lab/phyluce>.



## References

- [1] Nelson, G., 1969 Gill arches and the phylogeny of fishes, with notes on the classification of vertebrates. *Bull. Am. Mus. Nat. Hist* **141**, 465–552.
- [2] Jessen, H., 1972 Schultergurtel und pectoralflosse bei actinopterygiern. *Fossils Strata* **1**, 1–101.
- [3] Olsen, P., 1984 The skull and pectoral girdle of the parasemionotid fish *Watsonulus eugnathoides* from the Early Triassic Sakamena Group of Madagascar, with comments on the relationships of the holostean fishes. *J. Vert. Paleontol.* **4**, 481–499.
- [4] Patterson, C., 1973 Interrelationships of holosteans. In *Interrelationships of fishes* (eds. P. Greenwood, R. Miles & C. Patterson), pp. 207–226. London: Academic Press.
- [5] Arratia, G., 2001 The sister-group of Teleostei: Consensus and disagreements. *J. Vert. Paleontol.* **21**, 767–773.
- [6] Li, C., Lu, G. & Orti, G., 2008 Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst. Biol* **57**, 519.
- [7] Inoue, J. G., Miya, M., Tsukamoto, K. & Nishida, M., 2003 Basal actinopterygian relationships: A mitogenomic perspective on the phylogeny of the "ancient fish". *Mol. Phylogenet. Evol.* **26**, 110–120.
- [8] Patterson, C. & Rosen, D., 1977 Review of ichthyodectiform and other Mesozoic teleost fishes and the theory and practice of classifying fossils. *Bull. Am. Mus. Nat. Hist.* **158**, 83–172.
- [9] Arratia, G., 2010 The clupecocephala re-visited: Analysis of characters and homologies. *Rev. Biol. Mar. Oceanog.* **45**, 635–657.

- [10] Arratia, G., 2004 Mesozoic halecostomes and the early radiation of teleosts. In *Mesozoic fishes 3* (eds. G. Arratia & A. Tintori), pp. 279–315. Munich, Germany: Verlag Dr. Friedrich Pfeil.
- [11] Cloutier, R. & Arratia, G., 2004 Early diversification of actinopterygians. In *Recent advances in the Origin and Early Radiation of Vertebrates* (eds. G. Arratia, M. V. H. Wilson & R. Cloutier), pp. 217–270. Munich, Germany: Verlag Dr. Friedrich Pfeil.
- [12] Le, H., Lecointre, G. & Perasso, R., 1993 A 28S rRNA-based phylogeny of the gnathostomes: First steps in the analysis of conflict and congruence with morphologically based cladograms. *Mol. Phylogenet. Evol.* **2**, 31–51.
- [13] Broughton, R. I., 2010 Phylogeny of teleosts based on mitochondrial genome sequences. In *Origin and Phylogenetic Interrelationships of Teleosts* (eds. J. Nelson, H. Schultz & M. Wilson), pp. 61–76. Munchen: Verlag.
- [14] Inoue, J. G., Miya, M., Venkatesh, B. & Nishida, M., 2005 The mitochondrial genome of Indonesian coelacanth *Latimeria menadoensis* (*Sarcopterygii: Coelacanthiformes*) and divergence time estimation between the two coelacanths. *Gene* **349**, 227–235.
- [15] Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A., Davis, M. P., Wainwright, P. C., Friedman, M. & Smith, W. L., 2012 Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13698–13703.
- [16] Zou, M., Guo, B., Tao, W., Arratia, G. & He, S., 2012 Integrating multi-origin expression data improves the resolution of deep phylogeny of ray-finned fish (Actinopterygii). *Sci. Rep.* **2**.
- [17] Faircloth, B., McCormack, J., Crawford, N., Harvey, M., Brumfield, R. & Glenn, T., In Press Ultraconserved elements anchor thousands of genetic markers for target enrichment spanning multiple evolutionary timescales. *Syst. Biol.* (doi:10.1093/sysbio/sys004).

- [18] McCormack, J., Faircloth, B., Crawford, N., Gowaty, P., Brumfield, R. & Glenn, T.,  
2012 DNA flanking ultraconserved elements provides novel phylogenetic markers and re-  
solves placental mammal phylogeny when combined with species tree analysis. *Genome*  
*Res.* **22**, 746–754. (doi:10.1101/gr.125864.111).
- [19] McCormack, J., Harvey, M., Faircloth, B., Crawford, N., Glenn, T. & Brumfield, R.,  
*In review* A phylogeny of birds based on over 1,500 loci collected by target enrichment  
and high-throughput sequencing. *PLoS ONE*.
- [20] Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K.  
& Glenn, T. C., 2012 More than 1000 ultraconserved elements provide evidence that  
turtles are the sister group of archosaurs. *Biol. Lett.* **8**, 783–786.
- [21] Grande, L., 2010 An empirical synthetic pattern study of gars (Lepisosteiformes) and  
closely related species, based mostly on skeletal anatomy: the resurrection of Holostei.  
*Copeia* pp. 1–863.
- [22] De Pinna, M., 1996 Teleostean monophyly. In *Interrelationships of Fishes* (eds. M. L.  
Stiassny, L. R. Parenti & G. D. Johnson), pp. 147–162. San Diego: Academic Press.
- [23] Zaragüeta-Bagils, R., Lavoue, S., Tillier, A., Bonillo, C. & Lecointre, G., 2002 As-  
sessment of otocephalan and protacanthopterygian concepts in the light of multiple  
molecular phylogenies. *C. R. Biol.* **325**, 1191–1207.
- [24] Harris, R., 2007 *Improved pairwise alignment of genomic DNA*. Ph.D. thesis, The  
Pennsylvania State University.
- [25] Tewhey, R., Nakano, M., Wang, X., Pabon-Pena, C., Novak, B., Giuffre, A., Lin, E.,  
Happe, S., Roberts, D., LeProust, E. *et al.*, 2009 Enrichment of sequencing targets from  
the human genome by solution hybridization. *Genome Biol.* **10**, R116.

- 434 [26] Faircloth, B. C. & Glenn, T. C., 2012 Not all sequence tags are created equal: Designing  
435 and validating sequence identification tags robust to indels. *PLoS ONE* **7**, e42543. (doi:  
436 10.1371/journal.pone.0042543).
- 437 [27] Zerbino, D. & Birney, E., 2008 Velvet: Algorithms for de novo short read assembly  
438 using de Bruijn graphs. *Genome Res.* **18**, 821–829.
- 439 [28] Katoh, K., Kuma, K., Toh, H. & Miyata, T., 2005 MAFFT version 5: Improvement in  
440 accuracy of multiple sequence alignment. *Nucleic Acids Res.* **33**, 511–518.
- 441 [29] Nylander, J., 2004 *MrAIC.pl*. Evolutionary Biology Centre, Uppsala University: Pro-  
442 gram distributed by the author.
- 443 [30] Ronquist, F. & Huelsenbeck, J. P., 2003 MrBayes 3: Bayesian phylogenetic inference  
444 under mixed models. *Bioinformatics* **19**, 1572–1574.
- 445 [31] Stamatakis, A. & Ott, M., 2008 Efficient computation of the phylogenetic likelihood  
446 function on multi-gene alignments and multi-core architectures. *Philos. Trans. R. Soc.*  
447 *Lond., B, Biol. Sci.* **363**, 3977–3984.
- 448 [32] Kubatko, L. & Degnan, J., 2007 Inconsistency of phylogenetic estimates from concate-  
449 nated data under coalescence. *Syst. Biol.* **56**, 17–24.
- 450 [33] Edwards, S., Liu, L. & Pearl, D., 2007 High-resolution species trees without concatena-  
451 tion. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5936–5941.
- 452 [34] Edwards, S., 2009 Is a new and general theory of molecular systematics emerging?  
453 *Evolution* **63**, 1–19.
- 454 [35] Bayzid, M. S. & Warnow, T., 2012 Estimating optimal species trees from incomplete  
455 gene trees under deep coalescence. *J. Comput. Biol.* **19**, 591–605.

- [36] Castillo-Ramrez, S., Liu, L., Pearl, D., Edwards, S., Knowles, L. & Kubatko, L., 2010 Bayesian estimation of species trees: A practical guide to optimal sampling and analysis. In *Estimating species trees: Practical and theoretical aspects*, pp. 15–33. Hoboken: Wiley-Blackwell.
- [37] Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O., 2010 New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321.
- [38] Liu, L., Yu, L., Kubatko, L., Pearl, D. K. & Edwards, S. V., 2009 Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.* **53**, 320–328.
- [39] Liu, L., Yu, L., Pearl, D. K. & Edwards, S. V., 2009 Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* **58**, 468–477.
- [40] Liu, L. & Yu, L., 2010 Phybase: an R package for species tree analysis. *Bioinformatics* **26**, 962–963.
- [41] Seo, T.-K., 2008 Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol. Biol. and Evol.* **25**, 960–971.
- [42] Hurley, I., Mueller, R., Dunn, K., Schmidt, E., Friedman, M., Ho, R., Prince, V., Yang, Z., Thomas, M. & Coates, M., 2007 A new time-scale for ray-finned fish evolution. *Proc. R. Soc. London, Ser. B* **274**, 489–498.
- [43] Santini, F., Harmon, L., Carnevale, G. & Alfaro, M., 2009 Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evol. Biol.* **9**, 194.
- [44] Drummond, A. & Rambaut, A., 2007 BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214.

- [45] Li, C., Lu, G. & Orti, G., 2008 Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. *Syst. Biol.* **57**, 519–539.
- [46] Inoue, J. G., Miya, M., Tsukamoto, K. & Nishida, M., 2003 Basal actinopterygian relationships: A mitogenomic perspective on the phylogeny of the “ancient fish”. *Mol. Phylogenet. Evol.* **26**, 110–120.
- [47] Venkatesh, B., Erdmann, M. & Brenner, S., 2001 Molecular synapomorphies resolve evolutionary relationships of extant jawed vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11382–11387.
- [48] Mabuchi, K., Miya, M., Azuma, Y. & Nishida, M., 2007 Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evol. Biol.* **7**, 10.
- [49] Wainwright, P., Smith, W. & Price, S., *In Press* The evolution of pharyngognathy: A phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. *Syst. Biol.* .
- [50] Smith, W. & Wheeler, W., 2006 Venom evolution widespread in fishes: A phylogenetic road map for the bioprospecting of piscine venoms. *J. Hered.* **97**, 206–217.
- [51] Smith, W. & Craig, M., 2007 Casting the percomorph net widely: The importance of broad taxonomic sampling in the search for the placement of serranid and percoid fishes. *Copeia* **1**, 35–55.
- [52] Dettai, A. & Lecointre, G., 2005 Further support for the clades obtained by multiple molecular phylogenies in the acanthomorph bush. *C. R. Biol.* **328**, 674–689.
- [53] Meynard, C. N., Mouillot, D., Mouquet, N. & Douzery, E. J. P., 2012 A phylogenetic perspective on the evolution of mediterranean teleost fishes. *PLoS ONE* **7**, e36443. (doi:10.1371/journal.pone.0036443).

- 503 [54] Near, T., Meylan, P. & Shaffer, H., 2005 Assessing concordance of fossil calibration  
504 points in molecular clock studies: An example using turtles. *Am. Nat.* **165**, 137–146.
- 505 [55] Johnson, G., Ida, H., Sakaue, J., Sado, T., Asahida, T. & Miya, M., 2012 A “living  
506 fossil” eel (Anguilliformes: Protanguillidae, fam. nov.) from an undersea cave in Palau.  
507 *Proc. R. Soc. Lond., B, Biol. Sci.* **279**, 934–943.
- 508 [56] Janvier, P., 1996 The dawn of the vertebrates: Characters versus common ascent in the  
509 rise of current vertebrate phylogenies. *Palaeontology* **39**, 259–287.
- 510 [57] Patterson, C., 1993 Osteichthyes: Teleostei. In *The Fossil Record 2* (ed. M. Benton),  
511 pp. 621–656. London: Chapman & Hall.
- 512 [58] Santini, F. & Tyler, J. C., 2003 A phylogeny of the families of fossil and extant  
513 tetraodontiform fishes (Acanthomorpha, Tetraodontiformes), Upper Cretaceous to re-  
514 cent. *Zool. J. Linn. Soc.* **139**, 565–617.

## Description of timetree calibration points

We assigned fossil calibrations to 14 nodes of the maximum likelihood phylogeny (Fig. 2) to enable divergence time estimation using BEAST v1.72. Below, we justify the bounds for each calibration point.

### MRCA of Actinopterygii

**Fig. 2, node 1; uniform calibration = 392-472 Ma**

We used the stegotrachelids fossils from the Givetian/Eifelian boundary ( $\approx 392$  Ma) used by [42] to date the minimum age of the root and assumed an upper bound of 472 Ma based upon the minimum age for the split between acanthodians and all other bony fishes [56].

### MRCA of Actinopteri

**Fig. 2, node 2, exponential calibration lower bound = 345 Ma, upper 95% = 392 Ma**

The oldest fossil belonging to this clade is *Cosmoptycius* from the Tournasian (Carboniferous, 359-345 Ma) [42]. We used the stegotrachelids fossils from the Givetian/Eifelian boundary ( $\approx 392$  Ma) used by [42] to date the crown actinopterygians to set the upper bound.

### MRCA of Holostei

**Fig. 2, node 3; exponential calibration lower bound = 284 Ma, upper 95% = 345 Ma**

The oldest fossil assigned to this clade is the neopterygian *Brachydegma caelatum*, from the Artkinsian (early Permian, 284 Ma) [42]. We used *Cosmoptychius* from the Tournasian (Carboniferous, 359-345 Ma) to set the upper bound [42].

### MRCA of Teleostei (Fig. 2, node 5; unconstrained)

The oldest crown teleost is *Anaethalion*, from the late Kimmeridgian (Jurassic, 152 Ma)



[43]. However, due to the appearance of fossils from several major teleost groups (e.g., the ostarioclupeomorph *Tischlingerichthys viohli*, the euteleost *Leptolepides sprattiformis*) in deposits of the same age it seems likely that teleosts arose considerably earlier. We left this node unconstrained in our analysis to obtain an estimate based upon other actionpterygian calibrations and the UCE sequence data.

#### MRCA of Osteoglossomorpha

**Fig. 2, node 13; lower bound = 112 Ma, upper 95% = 225 Ma**

The oldest taxon assigned to our osteoglossomorph clade (which only includes a subset of the major osteoglossomorph lineages) is the arapamid *Laeliichthys ancestralis* from the Aptian Areado Formation of Brazil (Cretaceous, 112 Ma) [57]. We used the Hiodontidae *Yanbiania wangqingica* from the Barremian (Cretaceous, 130 Ma) to set the upper boundary [43].

#### MRCA of Ostarioclupeomorpha

**Fig. 2, node 8; lower bound = 149 Ma, upper 95% = 225 Ma**

The oldest taxon assigned to this clade is *Tischlingerichthys viohli*, from the upper Tithonian (Jurassic, 149 Ma) [43]. We used the stem teleost *Pholidophorettes salvus* (Pholidophoridae), from the early Carnian/Julian (Triassic, 228-225 Ma) to set the upper bound [43].

#### MRCA of Euteleostei

**Fig. 2, node 9, lower bound = 152 Ma, upper 95% = 225 Ma**

The oldest taxon assigned to this clade is *Leptolepides sprattiformis* (Orthogoniklei-thridae) from the late Kimmeridgian (Jurassic, 152 Ma) [43]. We used the stem teleost *Pholidophorettes salvus* (Pholidophoridae), from the early Carnian/Julian (Triassic, 228-225 Ma) to set the upper bound [43].

#### MRCA of Elopomorpha

**Fig. 2, node 10, lower bound = 135 Ma, upper 95% = 225 Ma**

The oldest taxon assigned to this clade is the albulid *Albuloideorum ventralis*, from the early Hauterivian (Jurassic/Cretaceous border, 135 Ma) [43]. The oldest stem elopomorph is *Anaethalion*, from the late Kimmeridgian (Jurassic, 152 Ma) [43].

#### Characiformes vs Cypriniformes

**Fig. 2, node 16, lower bound = 100 Ma, upper 95% = 149 Ma**

The oldest taxon assigned to this clade is the characiform *Santanichthys diasii* from the Albian (Cretaceous, 112-100 Ma) [43]. We used *Tischlingerichthys vlohli*, from the upper Tithonian (Jurassic, 149 Ma) to set the upper bound [43]. Our prior assumed 100 My as the minimum age, and 149 Ma for the upper bound.

#### Salmoniformes vs Esociformes

**Fig. 2, node 12, lower bound = 125 Ma, upper 95% = 152 Ma**

The oldest taxon assigned to this clade is *Helgolandichthys schmidi*, from the early Aptian (early Cretaceous, 125 Ma) [43]. We used *Leptolepides sprattiformis* (Orthogonikleithridae) from the late Kimmeridgian (Jurassic, 152 Ma) to set the upper bound [43].

#### MRCA of Ctenosquamata

**Fig. 2, node 11, lower bound = 122 Ma, upper 95% = 152 Ma**

The otoliths assigned to *Acanthomorphorum forcallensis* from the Aptian (Early Cretaceous, 124-122 Ma) represents the oldest fossils assigned to this clade [43]. We used *Leptolepides sprattiformis* (Orthogonikleithridae) from the late Kimmeridgian (Jurassic, 152 Ma) to set the upper bound [43].

#### MRCA of Acanthopterygii

**Fig. 2, node 15, lower bound = 99 Ma, upper 95% = 122 Ma**

The oldest taxa assigned to this clade are various Beryciformes (e.g., *Hoplopteryx*, *Trachichthyoides*) from the Cenomanian (Late Cretaceous, 99 Ma) [43]. We used the

fossil otoliths assigned to "*Acanthomorphorum*" *forcallensis* from the Aptian (Early Cretaceous, 124-122 Ma) to set the upper bound [43].

### ***Gasterosteus* vs *Taenianotus***

**Fig. 2, node 21, lower bound = 85 Ma, upper 95% = 122 Ma** We used the oldest gasterosteiform, *Gasterorhamphus zuppichinii* from the Santonian (84-85 Ma) to date the minimum age of this split [57]. We used the fossil otoliths assigned to "*Acanthomorphorum*" *forcallensis* from the Aptian (Early Cretaceous, 124-122 Ma) to set the upper bound [43].

### **Lophiiforms vs tetraodontiforms**

**Fig. 2, node 20, lower bound = 85 Ma, upper 95% = 122 Ma**

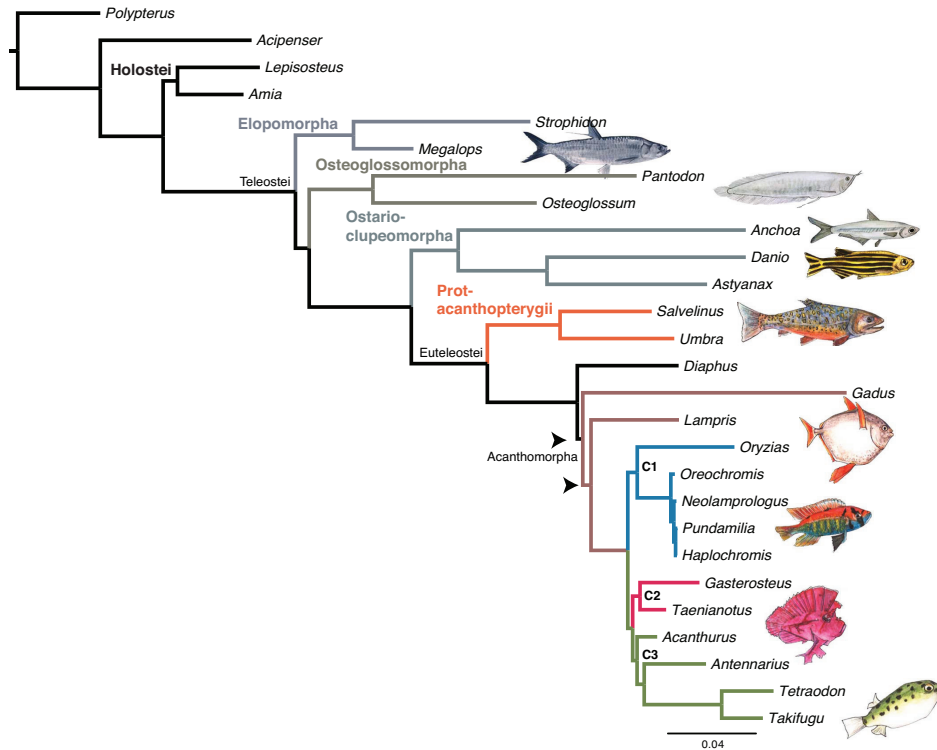
The oldest taxon assigned to this clade is the stem tetraodontiform *Cretatriacanthus guidottii* from the Santonian of Nardo (Italy). We chose this taxon to date the minimum age rather than other, older, stem tetraodontiformes because preliminary re-examination of the relationships of extant and fossil tetraodontiforms (Santini, unpublished) casts doubt on their phylogenetic affinities. We used the fossil otoliths assigned to "*Acanthomorphorum*" *forcallensis* from the Aptian (Early Cretaceous, 124-122 Ma) to set the upper bound [43].

### **MRCA of Tetraodontidae**

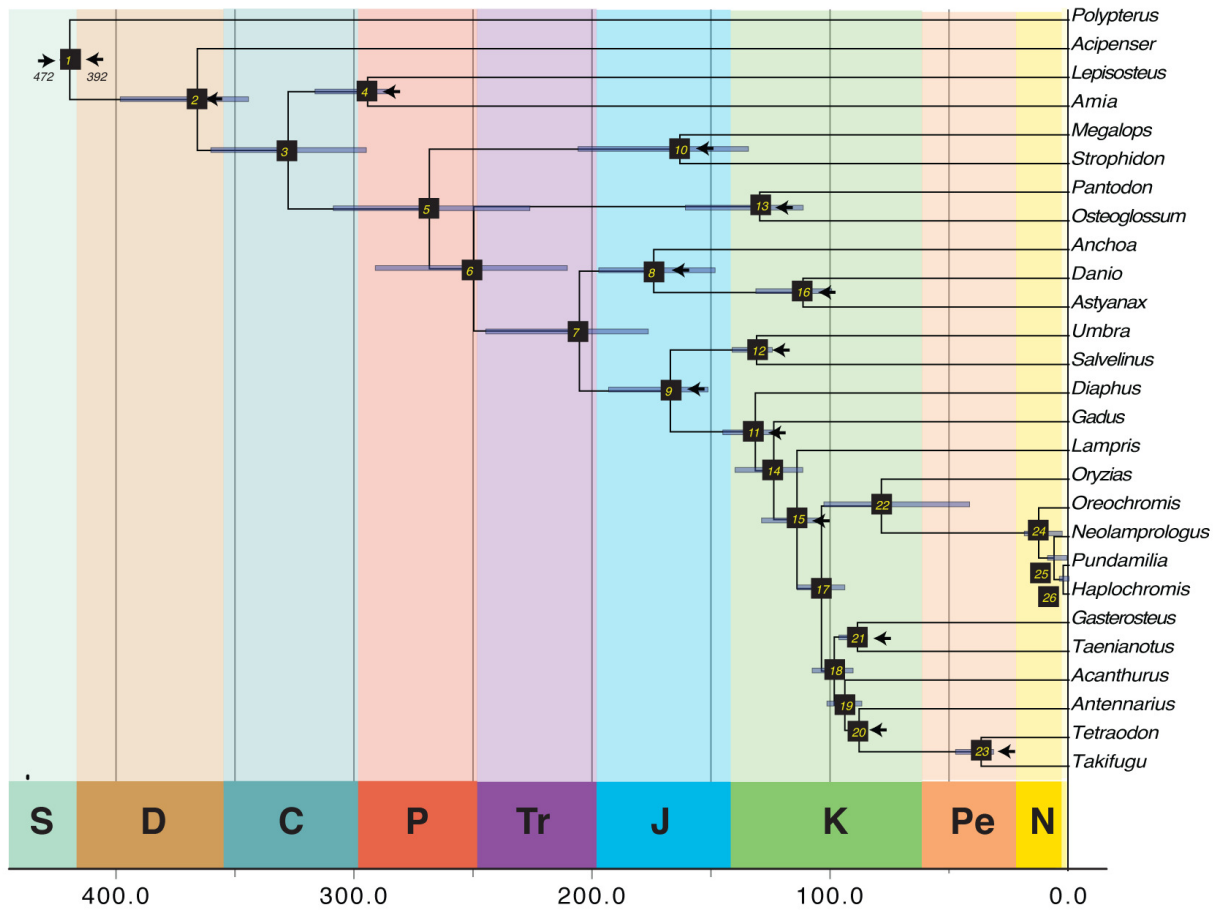
**Fig. 2, node 23, lower bound = 32 Ma, upper 95% = 50 Ma**

The oldest taxon assigned to this clade is *Archaeotetraodon winterbottomi* from the Oligocene of Caucasus (32-35 Ma) [43]. We used the stem tetraodontid *Eotetraodon pygmaeus* from the Ypresian (middle Eocene, 50 Ma) to set the upper bound [58].

## Figure Legends

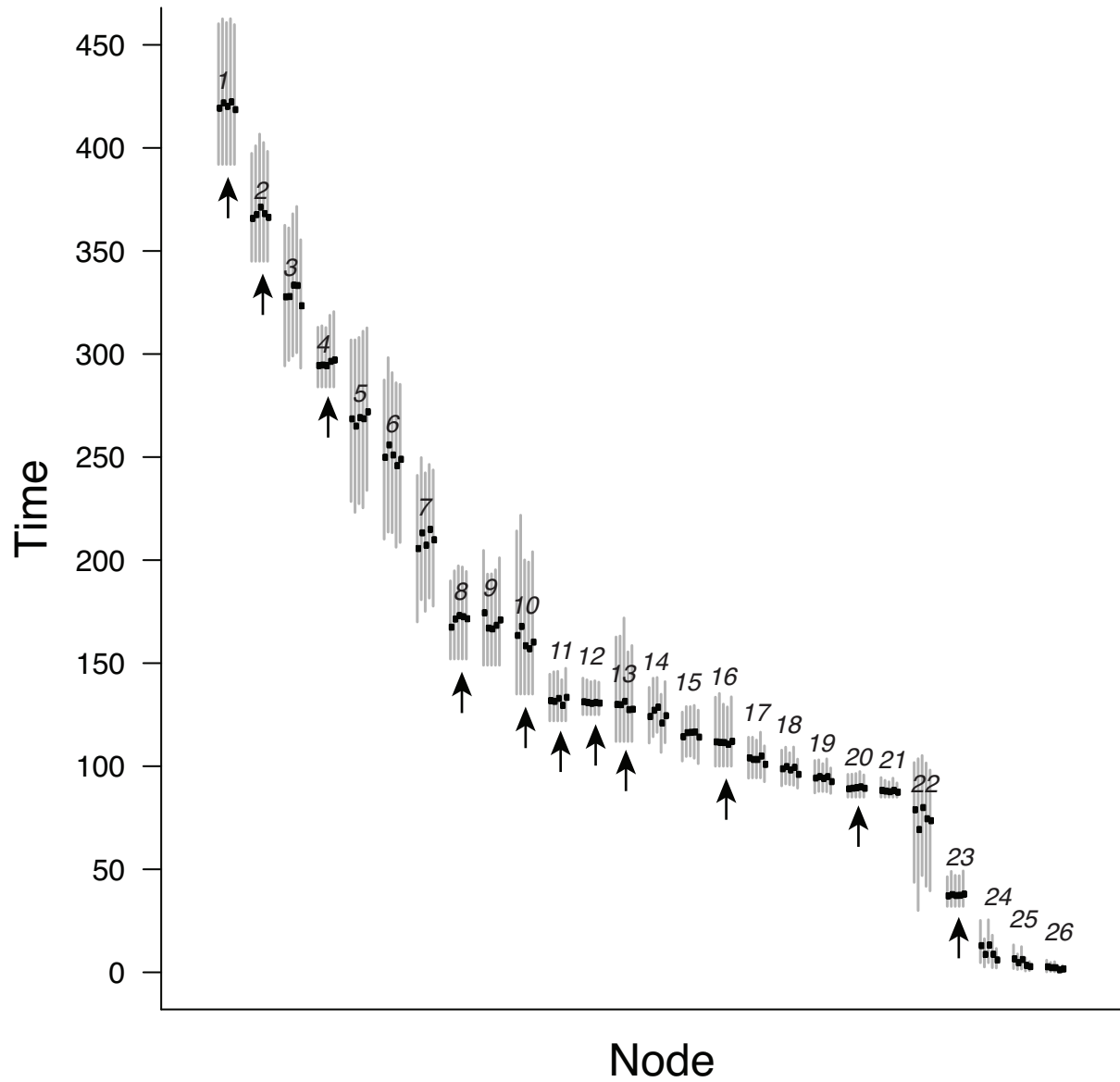


**Figure 1. Maximum likelihood phylogram of ray-finned fish relationships based upon UCE sequences.** All nodes except for two (indicated by arrows) supported by bootstrap proportions and Bayesian posterior probabilities  $> 0.99$ . Our analysis supports a monophyletic Holostei and reveals the elopomorphs to be the earliest diverging lineage of teleosts. C1, C2, and C3 indicate clades within acanthomorphs consistent with other recent molecular studies (see Discussion).

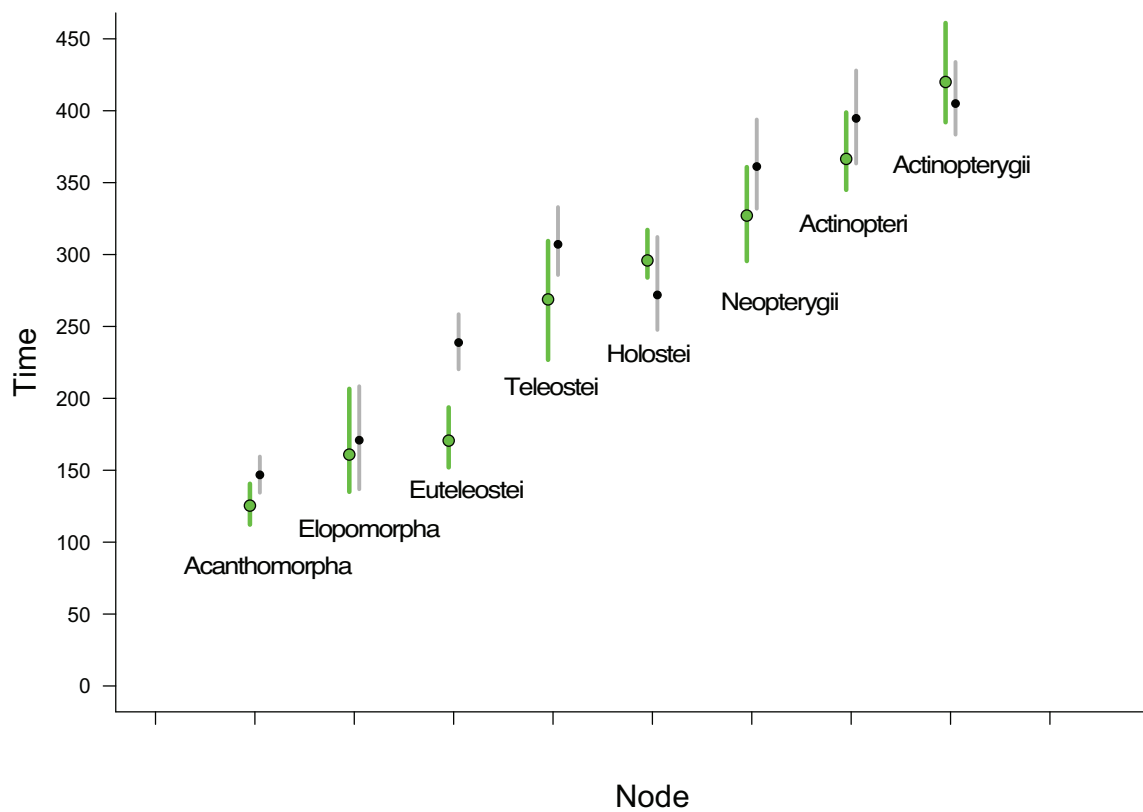


**Figure 2. UCE-based timescale of ray-finned fish diversification.** Arrows indicate fossil-calibrated nodes (described in Appendix 1). Numbers refer to splits in Fig. 1.





**Figure S2.** Mean and 95% credible interval of divergence time estimates for subsamples of the UCE data set. Node numbers refer to labels in Fig. 2. Arrows indicate fossil-constrained nodes. For each node, we show the mean and 95% credible interval of five divergence time analyses.



**Figure S3.** Comparison of ages derived from UCE data and nine nuclear genes [15]. UCE ages are in green. Circles indicate mean age and error bars indicate 95% credible interval.



# Tables

**Table 1.** Divergence times. Node numbers refer to Fig. 2. We averaged mean and 95% credible interval estimates (in Ma) across five subsampled divergence time analyses (see text).

Node	Clade	Split	Age	Min	Max
1	Actinopterygii	<i>Polypterus vs Takifugu</i>	420.0	392.0	461.0
2	Actinopteri	<i>Acipenser vs Takifugu</i>	366.5	345.0	398.8
3	Neopterygii	<i>Lepisosteus vs Takifugu</i>	327.1	295.5	360.8
4	Holostei	<i>Lepisosteus vs Amia</i>	295.9	284.0	317.1
5	Teleostei	<i>Megalops vs Takifugu</i>	268.8	226.8	309.4
6	Osteoglossocephala	<i>Pantodon vs Takifugu</i>	251.7	211.1	291.7
7	Clupeocephala	<i>Anchoa vs Takifugu</i>	210.9	177.1	245.3
8	Ostarioclupeomorpha	<i>Anchoa vs Astyanax</i>	169.5	149.0	197.8
9	Euteleostei	<i>Umbra vs Takifugu</i>	170.6	152.0	193.7
10	Elopomorpha	<i>Megalops vs Strophidon</i>	160.9	135.0	206.6
11	Ctenosquamata	<i>Diaphus vs Takifugu</i>	132.1	122.0	145.8
12	protacanthopterygii	<i>Umbra vs Salvelinus</i>	130.9	125.0	141.8
13	Osteoglossiformes	<i>Pantodon vs Osteoglossum</i>	128.9	112.0	161.5
14	Acanthomorphs	<i>Gadus vs Takifugu</i>	125.4	112.2	140.6
15	–	<i>Lampris vs Takifugu</i>	116.4	104.0	129.5
16	Ostariophysans	<i>Danio vs Astyanax</i>	111.4	100.0	131.9
17	–	<i>Oryzias vs Takifugu</i>	104.1	94.5	114.6
18	–	<i>Gasterosteus vs Takifugu</i>	99.1	91.0	108.2
19	–	<i>Acanthurus vs Takifugu</i>	94.1	87.4	102.0
20	–	<i>Gasterosteus vs Taenianotus</i>	89.7	85.0	97.0
21	–	<i>Antennarius vs Takifugu</i>	87.9	85.0	93.1
22	–	<i>Oryzias vs Haplochromis</i>	76.7	42.1	103.3
23	Tetraodontidae	<i>Tetraodon vs Takifugu</i>	37.5	32.0	48.0
24	Cichlidae	<i>Oreochromis vs Haplochromis</i>	9.9	3.2	19.2
25	–	<i>Neolamprologus vs Haplochromis</i>	4.7	1.3	9.4
26	–	<i>Pundamilia vs Haplochromis</i>	2.0	0.3	4.4

**Table 2.** Sequence read and assembly statistics for fish species used in this study.

Scientific name	Common name	Number of trimmed reads	Contigs assembled	Reads in contigs	UCE contigs	Reads in UCE contigs	Avg. size	Avg. coverage	Contigs on target	Reads on target
<i>Umbra limi</i>	central mudminnow	2,727,071	1109	740,079	409	564,715	508.8	267.4	0.37	0.21
<i>Diaphus theta</i>	California headlightfish	2,626,413	584	688,635	401	604,295	502.4	299.1	0.69	0.23
<i>Antennarius striatus</i>	striated frogfish	3,724,320	474	2,462,193	418	2,310,186	649.7	850.2	0.88	0.62
<i>Megalops</i> sp.	tarpon	2,771,805	786	650,577	247	231,314	485.4	191.5	0.31	0.08
<i>Astyanax fasciatus</i>	banded astyanax	2,731,668	543	1,444,767	355	1,211,903	526.2	657.2	0.65	0.44
<i>Acanthurus japonicus</i>	Japan surgeonfish	2,017,174	613	1,242,932	454	1,125,871	600.8	405.9	0.74	0.56
<i>Amia calva</i>	bowfin	2,619,643	562	1,608,614	366	1,368,091	578.9	646	0.65	0.52
<i>Lampris guttatus</i>	opah	2,472,439	486	1,350,852	418	1,237,650	568.7	520.2	0.86	0.50
<i>Acipenser fulvescens</i>	lake sturgeon	3,083,152	577	1,129,829	167	467,414	426.9	665.4	0.29	0.15
<i>Anchoa compressa</i>	deep body anchovy	2,617,717	533	783,323	287	625,862	448.6	479.2	0.54	0.24
<i>Danio rerio</i>	zebrafish	2,777,132	518	1,367,065	382	1,166,020	463.4	657.1	0.74	0.42
<i>Polypterus senegalus</i>	gray bichir	3,206,418	576	873,104	294	726,100	557.6	440	0.51	0.23
<i>Pantodon buchholzi</i>	freshwater butterflyfish	3,329,691	466	2,058,929	272	1,399,286	550.4	930.5	0.58	0.42
<i>Strophodon sathete</i>	slender giant moray	3,159,269	1007	448,390	277	246,758	510.6	172.9	0.28	0.08
<i>Osteoglossum bicirrhosum</i>	silver arawana	2,735,138	643	1,565,346	276	813,175	467	623.9	0.43	0.30
<i>Salvelinus fontinalis</i>	brook trout	2,466,696	1118	688,684	166	161,214	408.7	234.8	0.15	0.07
<i>Taenianotus triacanthus</i>	leaf scorpionfish	3,245,453	712	1,423,244	447	1,252,564	652.5	431.4	0.63	0.39